UNCLASSIFIED

Defense Technical Information Center
Compilation Part Notice

ADP010541

TITLE: Image Discrimination Models for Object
Detection in Natural Backgrounds

DISTRIBUTION: Approved for public release, distribution unlimited

This paper is part of the following report:

TITLE: Search and Target Acquisition

To order the complete compilation report, use: ADA388367

The component part is provided here to allow users access to individually authored sections of proceedings, annals, symposia, ect. However, the component should be considered within the context of the overall compilation report and not as a stand-alone technical report.

The following component part numbers comprise the compilation report:
ADP010531 thru ADP010556

UNCLASSIFIED

Image Discrimination Models for Object Detection in Natural Backgrounds

A. J. Ahumada, Jr.
NASA Ames Research Center
Mail Stop 262-2
Moffett Field, CA 94035
U.S.A.
E-mail: aahumada@mail.arc.nasa.gov

1. SUMMARY

This paper reviews work accomplished and in progress at NASA Ames relating to visual target detection. The focus is on image discrimination models, starting with Watson's pioneering development of a simple spatial model and progressing through this model's descendents and extensions. The application of image discrimination models to target detection will be described and results reviewed for Rohaly's vehicle target data and the Search 2 data. The paper concludes with a description of work we have done to model the process by which observers learn target templates and methods for elucidating those templates.

Keywords: target detection, image discrimination models, video quality metrics, target template learning, response correlation images, Cortex Transform, Discrete Cosine Transform, Minkowski summation

2. INTRODUCTION

The vision research laboratory at NASA Ames Research Center has developed a number of image discrimination and video discrimination models. Although these models are not themselves models for the search and detection of targets in complex scenes, they can be used to estimate the visibility of a target in a fixed, unchanging background. For some applications, this task may be a useful simulation of the detection task. Also the visual representation components of the models may be taken out and incorporated in more complete models of the search and detection situation.

An image discrimination model takes as input a pair of images and gives as output a number relating to the probability that the observer will be able to discriminate the difference between the two images. In our models each image is converted to a visual representation and then the difference between the two visual representations is aggregated using a Minkowski summation index. Differences among our models mainly result from different visual system representations.

3. MODEL REVIEW

3.1. Watson' s Simple Spatial Model

The first true image discrimination model was developed by Watson [1]. The basic element of the model was the linear sensor element with a Gabor receptive field similar to that of a simple cell in primary visual cortex. These cells were assumed to occur in quadrature pairs and to be arranged in layers of units that were self-similar, but spaced 1 octave apart in spatial frequency. Because the sampling was approximately adequate to represent all the pictures in the image, when euclidean distance was used as the summation exponent and the model was space-invariant, its predictions were the same as any single linear filter model with the same contrast sensitivity function.

3.2. Watson' s Cortex Transform Model

The next major advance in image discrimination models was Watson' Cortex Transform model [2]. The basic elements of this model are still linear orientation selective filters, but they are computed by means of a filtering scheme cleverly designed to be a pyramid transform. The transform domain amplitude is quantized according to a nonlinear just-noticeable-difference scale to provide automatic image compression [3]. This nonlinear scale provides a masking mechanism. If the background image has raised the amplitude level of a transform coefficient before a signal is added, the signal amplitude must be larger to cause a just-noticeable increase in the coefficient amplitude. The image quality metrics of Daly [4] and Lubin [5] are close decendents of this model. Watson also developed a version of the model that is much more computationally efficient by using the Discrete Cosine Transform as a crude approximation of the Cortex Transform [6].

3.3. Masking from other "cortical units"

Foley [7] has shown that not all masking can be explained by the nonlinear response of simulated cortical units using psychophysical measurements of Gabor targets masked by gratings of different frequencies and orientations. Watson and Solomon [8] developed an image discrimination model where units neighboring in space, spatial frequency, and orientation contribute to a divisive inhibition of each other. This model returned to Gabor-shaped receptive fields for the original linear filters, taking advantage of the increased speed of new computers and ignores image reconstruction from the visual representation.

3.4. Simplified Models

The increased complexity of the models with between-unit masking lead us to try models that used a simple global RMS contrast to provide the masking [9]. Although this model is easily shown to be wrong in detail because it lacks selectivity in position and spatial frequency, it can provide surprisingly good approximations to standard masking results. A slightly more complex version of the model was contructed to allow for background images that are not constant in luminance and RMS contrast [10].

3.5. Image Sequence Discrimination Models

A sophisticated detection model will base its visual representation on the spatio-temporal retinal signal. Our labs have developed two discrimination models for video sequences. One is a temporal extension of the DCT-based image model [11]. The other is an extension of the simplified model for non-homogenous backgrounds [12]. Both use recursive filtering in the time domain. The second model [11] keeps separate representations for a "parvo" channel (high spatial resolution, low temporal resolution) and a "magno" channel (medium spatial resolution, high temporal resolution).

4. DISCRIMINATION MODEL APPLICATION

4.1. Previous Results

Rohaly, Ahumada, and Watson have compared several discrimination models or metrics on their ability to predict target detection performance in natural backgrounds [13]. The target detection task had several simplifications from realistic

Paper presented at the RTO SCI Workshop on "Search and Target Acquisition", held in Utrecht, The Netherlands, 21-23 June 1999, and published in RTO MP-45.

detection tasks that made it suitable for the discrimination models. There was no search component to the task, the target if present was in the center of the image. For each target image a matched background image was made by replacing the target with a plausible section of background carefully blended with the original background. A discrimination experiment was run with the same monochrome, lower-resolution images that were given to the discrimination models. It showed that the detectability of the targets in the detection experiment, with different targets intermixed, was closely correlated to the discriminability of the reduced images with each target/no-target pair considered separately. Because of this, any visible difference between the images could be used as a basis for the detection.

Six models were tested. The first three were: 1) the difference between the images in the digital domain, 2) the difference between the images in the luminance contrast domain, contrast sensitivity filtered, 3) the Cortex Transform model (with a nonlinear amplitude transformation to provide within-unit masking). The next three were those model outputs normalized by a global contrast measure. For 4) the global contrast measure was simply the variance of the background image digital values. For 5) and 6) the contrast measure was the RMS contrast of the background luminance contrast image filtered with the contrast sensitivity function of the model.

The results were easy to remember. The order of the models, best to worse, with < meaning significantly better and = meaning approximately the same, was 4=5=6<3<1=2. Basically, the key to good model performance was masking. The two measures with no masking were the worst, the Cortex Transform model with within-unit masking was better, and any model with a global masking index was even better.

These results, and the results of some fixed noise masking studies with simulated airplane targets on runways, were the impetus for the development of the simplified model for nonhomogeneous backgrounds [10].

4.2. Search 2 Results

The single filter model with global masking was presented with small cut-outs of gray scale versions of the 44 Search 2 target images, together with matching versions with the targets removed [14].

The first step in the model calculations was to convert the images from digital gray scale, g, to luminance, Y, using the equation:

$$Y = 64.32((g-18)/(109.22+g-18))^{2.3} .$$

Next the images were converted to luminance contrast, C, using L0, the mean luminance of the image with the target removed.

$$C = Y/L0 - 1 .$$

The images were then filtered with a Difference-of-Gaussians contrast sensitivity filter. The center Gaussian had a spatial spread (1/e half width) of 2 arc minutes, the surround spread was 16 arc minutes, and the ratio of the surround volume to the center volume was 0.685. The filter was normalized to have a peak gain of unity in the frequency domain. The discriminability d' of the images is then estimated by the Euclidian distance between the filtered images, d, normalized by the Root-Mean-Square contrast of the filtered background image, c0.

$$d' = s\, d\, (1 + (c0/c2)^2)^{0.5} .$$

The parameter c2 is the masking threshold in contrast units and was set to 0.05. The sensitivity parameter s was set to give a contrast sensitivity of 114 for a filtered signal with constant unit contrast over an area of one square degree.

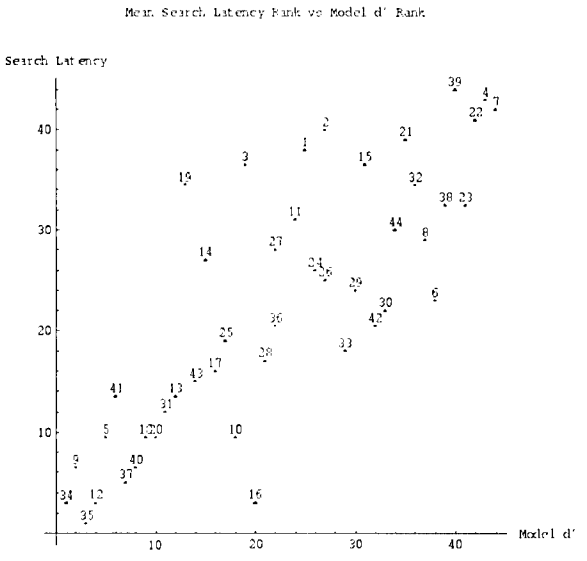

Figure 1

Figure 1 shows the ranks of the mean search latencies for the 44 signal images as a function the ranks of the model d's. The correlation coefficient for these ranks is 0.807, showing that this simple masking model predicts much of the variation in the search latencies.

This good performance is seen despite the fact that the model does not take into account (1) color differences, (2) target position, (3) object contours, or (4) texture differences.

The model does define and effectively combine (1) target size, (2) target contrast, and (3) background contrast variability.

It should be repeated that this is not a model of target detection. The hard part of the detection process was accomplished by the processes of limiting the image to the target region and replacing the target by a simulated background. For its simplicity, however, the model does a good job of capturing target visibility.

Two directions of improvement of this model are suggested. One is the addition of color. The other is a model that includes eccentricity. The model may have been improved by the poor color discrimination of the periphery. Also, in previous work with central targets we always found that a Minkowski distance with an exponent of 4 fit better than Euclidean distance, but for this data the Euclidean fit is better. This also may be a feature of peripheral search rather than foveal detection.

5. TARGET TEMPLATES

5.1. Fixed Noise vs. Random Noise Masking

Because we were interested in predicting detection performance in the presence of random noise, we have collected data comparing the relative effectiveness of random noise maskers and the fixed noise maskers that out image discrimination models are designed to predict [14]. A traditional signal detection approach to this problem predicts that random noise will be a stronger masker and that the difference should essentially depend on the ratio of the variability in the internal detection measure due to variability in the external noise to the internal variability in the measure when the noise is held constant. This internal variability was measured by Burgess and Colborne [17] for visual detection in noise by using the same noise on both intervals of a two-interval-forced-choice experiment. We call this method the twin noise method.

The interesting result comes from the comparison of twin noise with fixed noise. The standard models predict similar performance, but we find that fixed noise masking can be much less than twin noise masking [17].

5.2. Template Learning Models

To explain why fixed noise masking can be much less than twin noise masking we have developed models for template learning [18]. The basic idea is that in the fixed noise situation the observer is learning a template in a less variable situation, reducing the internal noise caused by variations in the template caused by the learning process itself. Another benefit of the fixed noise situation is that the template incorporates the fixed noise and reduces the spatial uncertainty in the detection process.

5.3. Template Identification Methods

A final research area that may be of interest to those trying to model target detection is our development of methods for identifying the features of the target that the observer is looking for. We add noise to the images in detection or discrimination tasks and correlate the noise pixels with the responses of the observers [19]. If the observer features are linear in the image pixel values, images of those features appear. If nonlinear features are used, the search is more tedious, but still possible [20].

6. ACKNOWLEDGEMENTS

I appreciate the assistance of Bettina L. Beard. This work was supported by NASA RTOP # 548-50-12.

7. REFERENCES

1. A. B. Watson (1983). Detection and recognition of simple spatial forms. In O. J. Braddick & A. C. Sleigh (Eds), *Physical and biological processing of images* (pp. 100-114). Berlin: Springer-Verlag.
2. A. B. Watson (1987a). The Cortex transform: rapid computation of simulated neural images. Computer Vision, Graphics, and Image Processing, **39**, 311-327.
3. A. B. Watson (1987b). Efficiency of an image code based on human vision. Journal of the Optical Society of America A, **4**, 2401-2417.
4. S. Daly (1993). The visible differences predictor: an algorithm for the assessment of image fidelity. In A. B. Watson (Ed.), *Digital Images and Human Vision* (pp. 179-206). Cambridge, Mass.: MIT Press.
5. J. Lubin (1993). The use of psychophysical data and models in the analysis of display system performance. In A. B. Watson (Ed.), *Digital Images and Human Vision* (pp. 163-178). Cambridge, Mass.: MIT Press.
6. A. B. Watson (1993). DCT quantization matrices visually optimized for individual images. B. Rogowitz and J. Allebach, eds., Human Vision, Visual Processing, and Digital Display IV, SPIE Proceedings, 1913, (SPIE: Bellingham, WA) pp. 202-216.
7. J. M. Foley (1994). Human luminance pattern-vision mechanisms: masking experiments require a new model. Journal of the Optical Society of America A, **11**, 1710-1719.
8. A. B. Watson & J. A. Solomon (1995). Contrast gain control model fits masking data. Investigative Ophthalmology and Visual Science, **36** (Suppl.), 438.
9. A. J. Ahumada, Jr. (1996). Simplified vision models for image quality assessment, J. Morreale, Society for Information Display International Symposium Digest of Technical Papers, Society for Information Display, Santa Ana, CA, 27, pp. 397-400.
10. A. J. Ahumada, Jr., B. L. Beard (1998). A simple vision model for inhomogeneous image quality assessment. J. Morreale, ed., Society for Information Display Digest of Technical Papers (Santa Ana, CA), 29, Paper 40.1.
11. A. B. Watson, J. Q. Hu, J. F. McGowan III,& J. B. Mulligan (1999). Design and performance of a digital video quality metric. In B. E. Rogowitz and T. N. Pappas, eds., Human Vision and Electronic Imaging IV, SPIE Proceedings, 3644, Paper 17.
12. A. J. Ahumada, Jr., B. L. Beard, R. Eriksson (1998). Spatio-temporal discrimination model predicts temporal masking functions. SPIE Proceedings, 3299, Paper 14, pp. 120-127.
13. A. M. Rohaly, A. J. Ahumada, Jr. & A. B. Watson (1997). Object detection in natural backgrounds predicted by discrimination performance and models. Vision Research, 37, pp. 3225-3235.
14. A. Toet, P. Bijl, F. L. Kooi, J. M. Valeton (1998). A high resolution image data set for testing search and detection models.TNO Report TM-98-A020, TNO Human Factors Research Institute, Soesterberg, The Netherlands.
15. A. J. Ahumada, Jr., B. L. Beard (1997). Image discrimination models predict detection in fixed but not random noise. Journal of the Optical Society of America A, 14, 2471-2476.
16. A. E. Burgess, B. Colborne (1988). Visual signal detection: IV. Observer inconsistency. Journal of the Optical Society of America A, 5, 617-628.
17. A. J. Ahumada, Jr., B. L. Beard (1997). Image discrimination models: detection in fixed and random noise. SPIE Proceedings, 3016, pp. 34-43.
18. B. L. Beard, A. J. Ahumada, Jr. (1999). Detection in fixed and random noise in foveal and parafoveal vision explained by template learning. Journal of the Optical Society of America A, 3, 755-763.
19. B. L. Beard, A. J. Ahumada, Jr. (1998). Technique to extract relevant image features for visual tasks. In B. E. Rogowitz and T. N. Pappas, eds., Human Vision and Electronic Imaging III, SPIE Proceedings, 3299, pp.79-85.
20. E. Barth, B. L. Beard, A. J. Ahumada, Jr. (1999). Nonlinear Features in Vernier Acuity. In B. E. Rogowitz and T. N. Pappas, eds., Human Vision and Electronic Imaging IV, SPIE Proceedings, 3644, Paper 8.